DOCUMENT RESUME

ED 228 276 TM 830 144

AUTHOR Burry, James
TITLE Evaluation and Documentation: Making Them Work Together.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO CSE-RP-2
PUB DATE 82
NOTE 38p.
PUB TYPE Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Documentation; Educational Planning; *Evaluation Methods; Formative Evaluation; *Information Needs; Interviews; Observation; *Program Descriptions; *Program Effectiveness; *Program Evaluation; Questionnaires; Records (Forms); Summative Evaluation

ABSTRACT

Program documentation is an important role for evaluation in both formative and summative modes. It should describe processes, materials, participants and social interactions which produce certain outcomes. Program evaluation should judge the effect of the program and depict effective practices, products and outcomes. To demonstrate program implementation and achievement, thorough documentation is required. Documentation methods should suit the needs of the individual program. These needs are likely to be influenced by audience; credibility; planning needs; decision makers at local, state, or other levels; time factors; and the nature of the program. At a minimum, documentation should provide information on program frame factors such as participants, resources and facilities, and program events and activities, including materials used and procedures. The principal approaches, used singly or in combination, gather information directly from participants by written descriptions, interviews, or questionnaires; program records; and direct observation of the program setting. Three points of view toward documentation validity are (1) consistency with the program plan; (2) plan-free documentation; and (3) theory-based documentation. These three points of view are illustrated with suggestions for practice and setting up the documentation system. (CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Gray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED228276

EVALUATION AND DOCUMENTATION: MAKING THEM WORK TOGETHER

James Burry

CSE Resource Paper No. 2

1982

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California, Los Angeles

TM 830 164

The work reported herein was supported in part under a grant from the National Institute of Education. However, the opinions expressed do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement should be inferred.

TABLE OF CONTENTS

INTRODUCTION 1

WHAT IS PROGRAM DOCUMENTATION? 4

How Extensive Should the Documentation Be? 5

DOCUMENTATION OF FRAME FACTORS 8

DOCUMENTATION OF PROGRAM ACTIVITIES 9

THREE POSSIBLE METHODS IN DOCUMENTATION 9

How Do You Make Decisions About What to Use? 10

INFORMATION GATHERED DIRECTLY FROM PROGRAM PARTICIPANTS 12

Participant Reports 12

Questionnaires 13

Interviews 13

EXAMINATION OF RECORDS 15

OBSERVATIONS 16

SUGGESTIONS FOR PRACTICE 23

What to Look At 23

From Which Point of View? 25

Setting up the Documentation 27

REFERENCES 34

EVALUATION AND DOCUMENTATION: MAKING THEM WORK TOGETHER

INTRODUCTION

When we evaluate a program, we should keep in mind the major purposes of that evaluation; that is, the role of the evaluation in the life of the program. One role of evaluation is formative; it serves to help and advise program planners and developers to describe and monitor program activities, assess the progress achieved, pick up potential problems, and find out what areas need improvement. Another major role of evaluation is summative; it is designed to provide a summary statement about the general effectiveness of the program: to describe it, judge achievement of its intended goals, pick up unanticipated outcomes, and possibly compare the program with similar ones.

Both the formative and the summative evaluator should be concerned with program description, with illustration of the dynamics of a program: its processes, materials, participants, and social interactions which produce certain outcomes. Describing a program in this way points to a third role for evaluation, one that is useful in both formative and summative modes. This is the role of program documentation.

It is important to note that I am discussing evaluation, rather than research. The distinction between evaluation and research is an important one, encompassing not only questions of

topic or technique, but also fundamental differences in how information is generated, the kinds of information generated, and their intended uses. Following Cronbach and Suppes (1969) who cite the need to distinguish decision-oriented from conclusion-oriented investigations, I categorize evaluation and documentation as decision-oriented study and research as conclusion-oriented study. Decision-oriented evaluation provides information on program results relevant for a particular program at a specific time. It differs from conclusion-oriented research, not so much in matters of topic of interest or techniques employed, but rather in the motivation for conducting the work, constraints imposed by the institutional setting in which the work is conducted, and intended use of the information generated. Thus, the first role of evaluation should be to generate information that will lead to decisions about educational questions and problems in a particular setting, while its secondary role is to generate information of use in other settings.

A great deal of what is called "evaluation" is in a conclusion- rather than a decision-oriented vein. Much of it deals with only the investigator's rather than the decision maker's information needs. Many evaluations lack in description and documentation; they describe outcomes without describing the processes and events that led to them: the interaction of instructional strategies and implementation techniques. Such evaluations offer a somewhat stilted picture of the program (Burry, 1979).

A good program evaluation, then, should not only judge the effect of the program but also depict its effective (or ineffective) practices, products, and outcomes. It should demonstrate how the program was implemented as well as what it achieved; this requires a thorough description and documentation of the program.

Implicit in the concepts of documentation and evaluation is the desire to discover those effective practices maintained in the parent site which may then be adopted at other sites. Such replication or adaptation of program practices depends strongly on the evaluator's description of the program's crucial features. These features include the activities, materials, people, and social factors that constitute the program and, in combination, led to a set of outcomes.

There is no single, universally appropriate way to gather sufficient documentation. The methods chosen should suit the documentation needs of the individual program and are likely to be influenced by:

1. The audiences--the people who will examine the results of the documentation either as a separate entity or as part of an evaluation.
2. The willingness of the audiences to accept the documentation as credible.
3. The needs of those who may plan to adapt/adopt the program in the future.
4. Other decision makers at local, state, regional, and federal levels.
5. The amount of time and help available to do the job.
6. The nature of the program documented.

WHAT IS PROGRAM DOCUMENTATION?

Documentation of a program is a compilation of records of the program's characteristics--participants, materials, activities, events, social factors--and a report (as part of the overall evaluation) of what the program looked like. It should describe all the crucial features of the program and be considered a normal product of program operation and evaluation. It can include a wide variety of materials such as attendance records, minutes of meetings, internal communications, materials and/or their descriptions, records formally set up to provide specific kinds of documentation, participant reports, interviews, and observations.

If positive results are found in a program, that program may be worth repeating, both to verify its effects and to give more people a chance to benefit from it. For these reasons, a simple statement of outcome is not sufficient in itself. Also needed is a thorough description which, in conjunction with statements of outcome, will document the dynamic interactions that helped define the program and its accomplishments. Similarly, if the results were negative, a thorough documentation might help determine what went wrong.

Documentation should be detailed enough to serve as a source for future planning. If another institution is impressed by a program and wants to implement it, such precise documentation will provide a source of planning for the new participants in their own setting.

Documentation, then, as part of evaluation, should help in data interpretation and in interrelation of process and outcome, a highly interpretative endeavor.

At a minimum, documentation should provide information on the following:

1. The frame factors of the program: the participants, resources, and facilities available.
2. The events and activities in the program: how materials were used and if they were used as intended; and the procedures and whether they were followed or modified to meet evolving needs.

These kinds of activities are time- and people-related events that occur during the program. It is more difficult to ascertain that proposed program activities are occurring than that a certain number and type of materials and participants are present. Also, some program activities are difficult to define and spot. For example, while we may easily document the presence of an aide in the classroom, how do we spot the presence of "reinforcement" or "acceptance of cultural diversity" in a classroom setting?

These intangible entities cannot be documented in the straightforward way that the presence of materials and numbers of participants can be verified. The choice of documentation techniques is therefore a difficult and sensitive business.

How extensive should the documentation be?

The documentation methods chosen will depend on the factors mentioned earlier--the audiences, the time and other resources available, and the nature of the program itself.

The Audiences. Documentation must be responsive both to internal program needs and to the needs and questions of different external audiences. The need for and amount of documentation materials collected on program activities depends primarily on the belief of the audience that the content of the documentation accurately portrays what happened in the program. For instance, an audience consisting largely of participants in the program would probably be in a good position to know if the documentation gives an accurate picture. On the other hand, external audiences such as funding agencies may be less inclined to accept the documentation at face value. If the audience consists of a group of people who may implement the program or parts of it at another site, then the conern in documenting will be to communicate the program's most crucial features to ensure that the documentation describes what occurred.

Time Available. Time determines the extent to which a program can be fully documented. Examination and interpretation of primary documents and records, construction of special documentation materials, interviewing, and observing are time-consuming. While many things in a program can conceivably be documented, it is wise to narrow down the list of these documentable activities to those of highest priority: those features that contribute to the overall effect of the program.

The Nature of the Program. This factor also determines the amount of documentation needed. The amount of documentation a program demands is often a function of that program's tangibility--how much it relies on observable materials relative to spe-

cific activities. For example, a program consisting of a series of printed, published materials is a fairly tangible program, while one that consists primarily of the implementation of processes (collaborative program planning in bilingual education or values clarification in multicultural education) may have to rely heavily on direct observation to document that these intangible processes indeed occur.

Further, many programs make strong prescriptions about the nature of activities that are to take place and how often they should happen. These demand more extensive documentation. A teacher training program, for example, where trainees are being instructed to use reinforcing techniques in a multicultural setting, will need extensive documentation. Direct observations, possibly by several observers, of how the trainees conduct themselves in this situation--in the classroom with children or when receiving training at a teacher training institution--will be needed. Observers will have to be trained and informed about exactly what "reinforcing" activities look like.

Another situation that calls for extensive documentation is one in which the planned program specifies certain minimal levels of behavior from the participants, for example, where aides or volunteers work with students at their desks 50 percent of the day. To determine whether this activity is as pervasive as intended, extensive documentation will be needed. Knowing how often and what percentage of time such activities have occurred will lead to better interpretation of process and outcome in the program.

A documentation effort will not always rely exclusively on the records or documents that naturally evolve during the course of the program. Depending on the nature of the program, we may want to influence the kind of documents that evolve so that they will meet our particular documentation needs.

DOCUMENTATION OF FRAME FACTORS

This documentation will look primarily at the **participants** and **materials** that are part of a program.

Some questions related to this part of the documentation are:

Personnel

What percentage of parents participated?

Are these parents typical of the whole school community or only of certain groups?

What special characteristics, training, abilities, or community affiliations are required for staff members?

Did the staff meet these requirements?

Materials

What new instructional materials were developed?

How did this development occur? By whom? When?

Where can a set of these materials be found for examination?

Participants

Who are the students in the program?

How old are they?

What is their linguistic background?

What are their SES levels?

How were students/classes selected?

How do they compare to other students in the district?

DOCUMENTATION OF PROGRAM ACTIVITIES

This documentation will look primarily at the events and activities of the program. People who may follow the program will need to know exactly what processes and activities occurred. If people want to form reasonable hypotheses about what critical program factors helped produce its outcomes, documentation will provide needed information. Further, discoveries of this kind will help form future policy decisions, thus giving evaluation a wider universe of relevance than would the single-project focus.

Where a project intends to implement a particular program and produce certain outcomes, documentation can demonstrate its accountability. This is particularly true in cases where a program aims toward outcomes far distant in the future or those that are difficult to recognize or measure accurately. In these cases, documentation may be the only basis for short-term accountability.

THREE POSSIBLE METHODS IN DOCUMENTATION

There are three principal approaches to documentation that can be used singly or in combination:

1. Gathering information directly from program participants: students, teachers, aides, parents, administrators, etc. These participants are asked to write descriptions of program activities. As an alternative, they may be interviewed or asked to complete a questionnaire.

2. Using the records kept over the course of the program (as well as influencing the generation of kinds of particular records that would be useful to specific documentation needs).

3. Conducting direct observations of the program operating in the classroom or elsewhere.

How do you make decisions about what to use?

What sort of documentation information would be most useful in a situation? Would the evaluator like to know how well the implemented program fit the original proposal? Or is the evaluator primarily interested in program outcome, regardless of the proposal? Does the evaluator want information on the program with respect to a particular model, philosophy, or theory of education? That is, does the program represent an example of this model, philosophy, or theory of education? These questions highlight three different points of view that might be taken toward program documentation. Each point of view has a slightly different purpose, but the purposes are not mutually exclusive.

1. Documentation of consistency with the program plan. Documentation from this point of view will involve collecting data to determine the extent to which the (crucial) activities as planned did occur; and if they did not occur, what happened instead.

2. Plan-free documentation. Sometimes program implementors lack consensus about what constitutes a program's crucial activities; sometimes these activities cannot be identified until the program begins. In this situation, and when plans are changed to meet evolving needs, plan-free

documentation will be useful. What actually occurred will be documented, regardless of what was intended. Documenting from the plan-free point of view, however, does risk omission from the documentation of various activities that some participants might feel were crucial. Its advantage is that a relatively unbiased picture of what the program actually looked like in operation will occur, and perhaps pervasive and crucial activities not originally conceived will be uncovered.

3. Theory-based documentation. This approach permits program documentation in terms of consistency with a theoretical prescription of how these activities should look; that is, whether the program is consistent with a specified educational model, theory, or philosophy. Documentation, in terms of theory, will provide valuable information for future planning, especially if the documentation is tied into evaluation measures of outcome. If the program has outcomes that are difficult to measure or that will occur far in the future, at least the documentation will demonstrate accountability for implementing activities that, by virtue of model or theory, are likely to bring about these outcomes.

While one particular documentation point of view may assume primary importance, the overall plan can call for the use of more than one viewpoint, if appropriate. In this way, a documentation effort might describe achievement of **intentions** (consistency), achievement of **unexpected** or **unplanned** effects (plan-

free), and the contribution of a particular educational philosophy (theory-based).

INFORMATION GATHERED DIRECTLY FROM FROGRAM PARTICIPANTS

Although the credibility of findings based on participant's reporting may be challenged by some, they can still contribute to a fairly accurate description of program activities. Gathering participants' reports will involve questioning participants especially in terms of, but not necessarily limited to, consistency with program plan or theory. There are three ways of accomplishing documentation via staff reports:

1. Participants can write a report of what they did and how these activities are consistent with those prescribed.
2. Questionnaires can be constructed for participants. This will involve forming a list of critical activities and framing questions about their occurrence.
3. Participants can be interviewed. The interview can either be **structured** (demanding the predesign of an interview questionnaire) or **unstructured** (requiring that participants describe how what they did was consistent with the program prescription). In the latter method, extensive notes must be recorded and transcribed.

Participant Reports

Eliciting participant reports is advantageous in that a special instrument can be quickly designed for them. An added advantage is its focus on participant perception of the program's important activities. These reports can be of two kinds: period-

ic reports throughout the course of the program or retrospective reports at its end.

Periodic reports will yield fairly accurate documentation. They do not make demands on memory like those required by end-of-program reports. On the other hand, end-of-program reports, and the resultant introspection, may provide a picture more tempered by reflection.

Questionnaires

Questionnaires must be written to elicit responses that are accurate descriptions of the program. There are two formats commonly used in writing questionnaires--closed (selected) or open (constructed) responses. Closed response scores easily and reports clearly, but tends to lock respondents into answers dicated by the items. Open response is time-consuming, but it allows respondents to express themselves freely and may provide a fuller picture. A questionnaire can--and frequently should--consist of both kinds of responses.

Interviews

Using interviews for program documentation has several advantages over participant reports and questionnaires. The interview is a personal, face-to-face contact. This offers flexibility in asking questions and provides an opportunity to obtain complete responses. If, for instance, it is discovered that a particular activity has not occurred, interviewing allows this question to be pursued further, probing for the respondent's opinion about why it was not used and discovering whether it was used in other class-

rooms. Such questions can provide more complete information than might the questionnaire or the participant report.

Another advantage of the interview is that the information gained will be more complete. Participant reports must rely upon the writer's recall of what occurred, and the writer's memory may be faulty. In the interview, the respondent can be cued and questioned about specific areas that might otherwise be overlooked. If carefully handled, the interview offers maximum access to information. It can, however, be time-consuming.

The form of the interview will depend on the frame of reference chosen for asking questions. Like the closed response questionnaire, an interview can be structured, based on specific questions in a prearranged order. Neither the questions nor their order are varied across interviewees. It can also be unstructured, consisting of a few general questions with predetermined probes designed to allow the respondent to amplify answers.

Self-report information from the program participants can be adequately covered through staff reports, questionnaires, and interviews. Yet, reports, questionnaires, and interviews of staff members, no matter how well carried out, rely upon the opinions of the staff members and the information they are willing to give. Questionnaires and interviews, further, may be poorly constructed, leading to biased information. In such cases, credibility can be increased by examining records or by direct observation.

EXAMINATION OF RECORDS

Records can provide systematic, tangible accounts of program occurrences. There are two things to consider about program documentation by record examination: (1) using existing program records, and (2) designing special recordkeeping systems for the documentation.

It is generally possible to use existing records kept over the course of a program already underway. In the absence of record systems specifically designed for documentation purposes, records normally kept for program administration or materials that result from the implementation of program activities will often yield a considerable amount of the information required.

An important advantage to using records is that they provide more credible information than staff reports. They also provide information on program activities without troubling the participants with requests for written reports or with time-consuming questionnaires and interviews. Records kept for the program, however, may not contain all the information required for an accurate program description. This is why records are more commonly used as sources of documentation for only **some** program activities. Unless recordkeeping is carefully planned for particular documentation needs, records seldom contain the whole story, and what they do contain may be inaccurate.

Another factor to consider is that, at times, there may be legal or ethical barriers to record examination. Problems over the legality of record examination often can be overcome by assur-

ances that data collected will be presented in summary form and will keep confidential the identities of individuals. Documentation by means of record examination can also be a time-consuming process: It takes time to obtain records, extract relevant information from them, and summarize the data (Stenhouse, 1977).

OBSERVATIONS

Any documentation can profit from on-site observation of the program in action. In many cases, this will merely involve the observer's presence during a program session. In other situations, observations of the program in action will be a major source of documentation data. Some people feel that observation is the method for performing program documentation. Observation, however, may itself interfere with the working of the program, causing anxiety and affecting the routine operation of the very activities to be observed.

Observation techniques can be informal or formal. **Informal observations** involve a casual look at the program by a few observers, followed by preparation of a report. They are quick and easy to achieve. **Formal** (or **systematic**) **observations** follow a more strictly controlled methodology, specifying observation times and recording methods.

Systematic observations should be considered when:

1. The program is guided by a highly specific plan or theory. The stronger the prescription, the more systematic the documentation should be.
2. The documentation and evaluation may be used to make generalizations about the causes of the program's outcomes.

3. The program's audiences may ask for strong supporting data.

In the absence of these kinds of conditions, systematic observation may not be necessary; but these conditions are likely to be there in some form or other.

The credibility attributed to systematic observation methods stems from the clear definition of these techniques; they prescribe what observers will watch and how they will record their observations.

The purpose of systematic observation is to provide evidence of whether a program was implemented as planned. But the observers do not judge; rather, they are human recording instruments who are transferring to paper a record of the activities that occur as the program operates.

Depending on the point of view from which the evaluation is being undertaken (plan-free, plan-based, theory-based, or some combination), observers may or may not be preinstructed about what behaviors to look for.

We can organize observation methods according to:

1. Whether or not the behaviors to be observed have been predefined.
2. How observers have been instructed to behave during and after the observation session--whether they sit and watch, write, make checks on a checklist, or reconstruct what happened after the observation period has ended.

Systematic observation is a technique where events and behaviors to be watched (or watched for) can be predefined. There are three methods available for conducting observations based on a predefined notion of how the program should look:

1. Checklists. The observer, with a list of defined behaviors, waits for them to occur. When they do, the observer checks that they have occurred at least once or tallies their frequency.
2. Coded behavior reports. The observer learns codes or symbols representing events and behaviors to be watched, keeps a running record of the sequence of occurrence of these events, and ignores events for which no symbols are provided.
3. Delayed reports. The observer is told which events/ behaviors to watch for and then observes the program in operation. When the prescribed period for the observation has ended, the observer fills out a checklist or questionnaire that records the occurrence and form of the behaviors that were watched, asks about the degree of frequency with which they occurred (rating scales), or uses a combination of recording methods.

These three observation techniques will be useful for obtaining slightly different types of data where different preconditions exist. Which observation method will most closely fit the documentation needs will depend on the information required. For instance, if frequency of occurrence is the primary concern, the checklist may be sufficient. If, however, the typical sequences of events must be described, the coded behavior will be needed.

Checklists will be useful when the evaluator wants to know whether or how often certain events or behaviors occurred. They

will not help document the form that they took. Where many behaviors/events are to be watched, it will be difficult to use a checklist. One of the two other methods--coded behavior or delayed report--will be more appropriate. A checklist can be used to tally frequency or duration of activities. The method for tallying frequency differs slightly from that for tallying duration. Duration demands some sort of time sampling, where observations are paced by the passing of short time intervals. Frequency tallying means tallying each occurrence of a discrete behavior or event.

Coded behavior records allow the evaluator to record the occurrence of several behaviors, though each of these behaviors must be precisely specified. Essential to the coded behavior record method is the assignment of symbols, or codes, to the behaviors observed. These codes will be used by observers to produce a quick, running account of the behaviors they observe. To achieve the desired account, observers will have to be trained to distinguish one behavior from another and to learn the symbol system. A good rule-of-thumb is to set up a code system comprising as few symbols as possible in order to ensure that observers will neither forget nor confuse them. If symbols can be combined to provide precise descriptions of behaviors, the repertoire of possible behavior descriptions will increase dramatically (depending on the number of permutations of combinable symbols). In other words, the number of symbols used should be as small as possible, but they should be combinable so that the number of possible precise behavior descriptions that can be made will still be extensive.

Coded behavior records are generally more time-consuming than simple checklists. Observers must be trained and tested for agreement on the meaning of symbols they will use to keep track of what they observe. Once these observations are taken, data must still be extracted from the observation records. Counting tallies or scoring checklists take minimal time, but coded behavior records produce strings of symbols. These symbol strings must be subjected to further interpretation for the information obtained to be suitable for reporting.

When should coded behavior records be collected? The coded behavior record will give more information about the form of behaviors--how they looked in operation--than will the checklist. Depending on how the categories are defined for coding, the evaluator will be able to detect whether activities happened in the way prescribed by the program plan or theory. Also, if the activities did not coincide with the plan, the evaluator will be able to describe what happened instead. The coded behavior record will also provide information about the form the activities took, showing patterns of events, activities, and behaviors.

Two common methods, tracking by event/behavior and tracking by time periods, are in general use for setting up a coded behavior report. When tracking by event, the observer marks a symbol each time it occurs, no matter how long it lasts. When tracking by a time period, the observer notes each time a prescribed time period passes. If the individual being observed performs an activity for a long time, the observer notes a symbol for the behavior displayed each time a period of particular length passes.

Tracking behavior is most useful for noting behavior or events that change often over the course of an observation, or when the event of interest has a clear beginning and end. Tracking by time periods works better where behaviors/events are prolonged. Obviously, it is possible to combine the two methods.

Delayed reports are checklists or questionnaires completed by the observer **after** a period of observation has ended. This "time unit" method, highly systematic, follows a clearly defined schedule. With questionnaire or checklist in hand, the observer watches **short** (a few minutes) segments of a participant's behavior, takes a predetermined number of minutes to fill out a long checklist or questionnaire, observes another time unit of the same or another person's behavior, then records, observes, etc. The observer can either record notes in the place of occurrence or withdraw to record the data, as when such activity would distract or produce anxiety in the participants. Observations made this systematically, particularly where reliability can be demonstrated between two or more observers, are highly credible. Of course, this highly regulated approach demands that the behaviors observed be describable via checklist or questionnaire; it also demands that the predetermined time segments be selected from periods of the day when the behaviors in question are likely to occur. Delayed reports are most useful where behaviors must be watched so carefully during the observation period that recording on-the-spot would interfere. It is also the method to choose when data from many participants must be gathered during an observation period. When the behaviors of many people are involved, observers can plan

to watch different people during each time segment: After recording behaviors of one individual for a given time segment, each observer can switch to a different person.

The delayed report methodology is less time-consuming than setting up and filling out a coding system. In reality, using the delayed report method is not much more complex than the writing of a questionnaire. There are, however, two procedures that must be followed in addition to developing a questionnaire:

1. The observers must be preinformed about what they are looking for by having them learn the questions on the questionnaire beforehand.

2. The observation session must be formalized as much as possible by controlling when it will occur, how long it will last, and who will be observed.

I suggest this close emphasis on **behaviors** not because I question an observers' ability to make inferences (though to an extent I do), but because inference in a recording instrument will inevitably affect the reliability of the observation method. Reliability is the degree to which one observer using an observation scheme gets the same results as another observer using the same scheme. It is a measure of the stability of a measuring method. Frequently, reports of an observation method are accompanied by an index of the instrument's reliability. The **correlation** between observations from each of two observers can serve as an index of reliability. Two or more observers can be relied upon to **see** the same thing as they watch the same behaviors, but they might not make the same **inference** from what they have seen. For the sake of agreement among observers, the inevitable infer-

ences they make must be kept as close as possible to the behaviors observed.

Where inference is inescapable, reliability can be increased by pretraining observers so that they agree about exactly which behaviors are referred to by each item on the recording instrument. A relatively high-inference report system that can demonstrate high reliability (a correlation of .80 or better) will be as useful and credible as any other.

As noted earlier, knowing which documentation technique(s) to use is difficult. In addition, deciding whether the documentation should stress program and plan consistency, be plan-free, illustrate theory, or take the form of some combination is also difficult. Further, some of the particular techniques described require more time and effort in their proper uses.

Careful documentation, however, adds to the evaluation component an element that fully describes program implementation and practices and those events and processes that interacted to achieve outcomes. This goes far beyond the typical process of statement of outcome and estimation of effect. It recognizes program uniqueness and documents what a particular program did to achieve its particular outcomes with a certain kind of student.

SUGGESTIONS FOR PRACTICE

What to Look At

Evaluators must become thoroughly familiar with a wide variety of program features. They must build into their information-gathering plan a variety of formative tasks to include examination

of program implementation, using resultant information to suggest areas of program improvement and careful documentation of the implemented features and their relationships. Evaluators must also improve their summative evaluation activities by estimating the contribution of specific program features to program outcomes. Evaluators must single out appropriate program features for examination in an evaluation that is both technically sound and sensitive to the needs and constraints of the program under examination.

These aims can be facilitated by examining the program plan or description; determining (in consultation with program participants) which program features need to be evaluated; deciding upon the means to evaluate these features; selecting methods to ensure ongoing documentation of their implementation, relationships, and cumulative effects; and considering appropriate means of reporting information to various audiences.

Features most needing evaluation and documentation will vary from program to program. Selection of these features and management of their evaluation/documentation can be facilitated through examining regulations governing the program, program plans and records, previously gathered evaluative information, and discussions with staff. Evaluative feedback on these features may enhance the information's utility for program staff. Such reporting, also, will be invaluable to others who may use the information to develop and implement a similar program.

From Which Point of View: Consistency, Plan-Free, Basis in Theory, or Combination?

Consistency with program plan. This point of view presupposes the existence of some sort of guiding plan or description. Collection of documentation data will determine the extent to which the crucial activities of the program occurred as planned; if they did not, the documentation should pick up what happened instead. This information, together with other evaluation data, is most useful to program staff who want formative evaluation information--information that will help them decide whether the program is working as intended; if not, why not; areas where the program can be modified and improved; and areas in which the program's operation is optimal but for which a written plan was poor or did not anticipate unforeseen events. This approach to documentation is also extremely useful in the summative evaluation of a program; it provides present program operators and potential second-generation program users with a careful description of implementation and the setting most conducive to its success.

Plan-free documentation. This approach to documentation avoids preconceptions about how program activities should look. It may be used in conjunction with the consistency approach. There are several situations in which plan-free documentation might be useful:

1. The program has no plan, nor is it based on a theory or model; it may be very appropriate in a program that is in its very early and evolving stages.

2. There is a plan, but it is either too vague or too grandiose to be practical for identifying documentation needs.

3. There is an apparent need for more naturalistic documentation that might be used in conjunction with a plan-consistency or theory approach to provide a fuller picture of the program.

Plan-free documentation, especially of complex programs, allows for uncovering important features or activities as implemented that might have gone unmentioned in the plan. This will often occur where there are strong social and political forces likely to affect the program. But this kind of documentation, as a sole source, may overlook activities crucial to program planners. The decision must be made whether the trade-off is likely to be favorable for producing a useful and credible documentation.

Theory-based documentation. This approach is useful for programs that attempt to implement a model of teaching, a theory of learning, or a philosophy of schooling. These phenomena may have identifying characteristics, prescriptions about the nature of activities in which the participants should engage. Proponents of each theory, philosophy, or model will contend that its particular activities will be the cause of the program's hoped-for outcomes. Documentation based on theory will then provide a check on whether these prescribed activities are indeed occurring and the extent to which they typify what happens to participants during program operation. This approach is also particularly appropriate for outcomes that are not directly measurable or that will not be realized until some future time, often after the program has ended. In this kind of situation, it is best used together with one of the two methods described above.

Setting Up the Documentation

The following suggestions are intended to help ensure a documentation system that will provide valid, reliable, and useful information. They amplify the three basic approaches to documentation outlined earlier.

1. Information gained from program participants. This can consist of reports, questionnaires, and interviews.

   a. **Participant reports** can be periodic and gathered throughout the course of the program, or retrospective and collected at its end.

      - If you rely on periodic reports, inform participants about the documentation effort, why you need this information, and how it will be useful to program participants as well as to the evaluation effort. List the reports you will need. List the participants involved and the dates on which you need reports, prepare instructions for the writers, and gather and interpret their reports.
      - If you rely on retrospective reports, decide when you will need them, and provide writers with sufficient lead time to meet your deadlines. Decide how many reports you will need, keeping in mind concerns with data reliability and credibility. Prepare instructions for the writers and gather and interpret their reports.

   b. **Questionnaires** can be closed- or open-response format. The steps in questionnaire development consist of:

- Preparing a program activities checklist. This checklist will refer to the program plan or theory, level of detail, frequency and duration of events, form that the events take, and degree of participant involvement.
- Considering question categories. Consider accuracy of questions in terms of occurrence, frequency, duration of activities as well as form and degree of involvement.
- Designing the questionnaire. This can take the form of closed response, open response, or a combination of the two.
- Critical examination of the questions: Are they straightforward? Do they imply a hidden agenda? Will they lead to biased answers?
- Testing and revising the questionnaire as necessary. Are the questions interpreted in the way you intended?
- Administering, analyzing, and interpreting results.

c. **Interviews** can be structured or unstructured. In writing questions for the structured interview:

- Prepare a program activities checklist (with the same kind of concerns followed for questionnaire development).
- Generate questions based on the list of activities; prepare alternate phrasing and probings to meet

nonexplicit responses. Again, the focus is on occurrence, frequency, duration, form, etc.

- Critically examine the questions; make sure the questions do not answer themselves.

In writing questions for the unstructured interview:

- Follow the same focus as above, but generating only a few general questions to be followed by probes to elicit detail.

For both forms of the interview:

- Assemble the interview form: Make sure questions are in logical order and that there is plenty of space to write answers.
- Write an introduction that sets the tone, states why you need the information, how you will use it, and describes data confidentiality.
- Rehearse the interview, looking for inconsistency in the logic of question sequencing and difficult or threateningly worded questions.

2. Examination of records. These can be record systems designed specifically for the program's documentation or they can consist of records that naturally evolve during the life of the program. Again, set up a program activities list reflecting plan, theory, or philosophy. Find out which program records will naturally be maintained. Determine which of these records will meet your documentation needs in terms of program events and their occurrence, duration, form, and participant involvement.

a. If the planned records do not entirely meet your documentation needs, discuss with staff/participants ways to incorporate the records you want.

b. Prepare a plan for collecting and extracting data: You will probably not be able to collect and transcribe data on every participant and event over the course of the program. What's more, to produce a detailed and credible documentation, you may not need to cover the entire spectrum of activities.

There are two possible record collection plans for reducing the amount of data:

a. Reduce the number of records to examine by selecting a sample (random, representative) of **participants and events** and base your documentation on these records only.

b. Reduce the number of records to examine by sampling **time units over the course of the program:** sample of days, weeks, or even months.

3. Observations. These can be informal or systematic. If you rely on informal observations:

a. Explain to the observers how the program should look or describe its underlying theory or philosophy.

b. Tell the observers what to look for.

c. Have each observer look at the program two or three times during its operation.

d. Have observers write reports describing what they saw, or give each observer a questionnaire, or interview them.

Systematic observations can consist of checklists, coded reports, or delayed reports.

Systematic observation consists of placing one or more observers into a real situation in the program. The observation is systematic because:

a. Observers enter at a preselected, typical time in the program's life.

b. They stay for a pre-appointed length of time.

c. They are usually guided in their observation by some form of pre-instruction.

The steps involved in carrying out systematic observations consist of the following:

a. Construct an activities checklist reflect plan, theory, and philosophy.

b. Observe the program in operation to get a perspective about what the observers are likely to encounter; this will help you with questions for the observation instrument.

c. Prepare an activities scenario to cover the events and behaviors that **should** take place as well as those that **should not.** It is important to get information on both.

d. Choose an observation method: checklist, coded behavior report, or delayed report. The method

chosen depends on the events you described in your scenarios as well as your documentation needs.

Checklist

Use a checklist when:

1. You want to measure how long a behavior lasted or how often it occurred.
2. You want to measure participant involvement and you feel that duration or frequency of behaviors of the participants will reflect involvement.
3. Your scenario depicts (or you wish to document) very few (six or fewer) crucial behaviors to be monitored.
4. The scenario depicts relatively few behaviors to be displayed in a fixed sequence.

Coded Behavior Report

Use a coded behavior report when:

1. You want to record the sequences in which behaviors occurred, including those sequences not prescribed.
2. You must document many (more than 6 and as many as 100) behaviors.
3. The amount of time available and your own expertise make it feasible to devise a code for recording behaviors.
4. You want to record as much as possible of what the observer sees.

Delayed Reports

Use the delayed report method when:

1. You feel you can get better data by sampling an individual's behaviors periodically, rather than recording whole episodes; for example, when you need data on many people working independently.
2. You are relatively short on time for training observers and processing data.

3. Behaviors must be watched so carefully that recording would interfere.

4. There is a possibility that recording would intimidate, fascinate, or otherwise disturb the people observed.

Regardless of which observation technique is selected, you must also decide how long each observation should last in order to get a good picture of the program. You will also have to decide **who** to observe and **when**, over the course of the program, to observe. It is better not to do all the observations over a limited time period, but rather to space them out over the program's beginning, middle, and end. Spreading out the observations in this way helps to detect changes in the program, increases representativeness of the documentation, and therefore helps increase its credibility as well as its interpretive power for the evaluation.

To the extent that it is necessary, observers will need to be trained in the use of the particular technique and the phenomena to be observed. This training is critical to the generation of valid and reliable documentation information.

Perhaps the best way to get a full and accurate picture of the documented program is to use multiple measures and data gathering methods that let you assemble converging data. To the extent that it is **appropriate** to the program and **feasible**, a combination of interviews, records, and observations can be used to generate information that supports or qualifies the picture of the program gained by each single approach.

REFERENCES

Burry, James. Evaluation in Bilingual Education. **Evaluation Comment**, 1979, 6(1), 1-14.

Cronbach, Lee J., and Suppes, P. (Eds.). **Research for Tomorrow's Schools: Disciplined Inquiry for Education.** New York: Macmillan, 1969.

Stenhouse, L. **Case Study as a Basis for Research in a Theoretical Contemporary History of Education.** Norwich, England: Centre for Applied Research in Education, University of East Anglia, 1977.